ELSEVIER
Stochastic Processes and their Applications 99 (2002) 137–157
stochastic processes and their applications
www.elsevier.com/locate/spa

Poisson limits for U-statistics

André R. Dabrowski[a,*], Herold G. Dehling[b], Thomas Mikosch[c], Olimjon Sharipov[d]

[a] *Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, PO Box 450 STNA, Ottawa, Ont., Canada K1N 6N5*
[b] *Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany*
[c] *Laboratory of Actuarial Mathematics, Institute of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark*
[d] *Department of Probability Theory, Institute of Mathematics, Uzbek Academy of Sciences, F. Khodjaev str. 29, Tashkent 700143, Uzbekistan*

Received 5 January 2001; received in revised form 6 September 2001; accepted 30 October 2001

Abstract

We study Poisson limits for U-statistics with non-negative kernels. The limit theory is derived from the Poisson convergence of suitable point processes of U-statistics structure. We apply these results to derive infinite variance stable limits for U-statistics with a regularly varying kernel and to determine the index of regular variation of the left tail of the kernel. The latter is known as correlation dimension. We use the point process convergence to study the asymptotic behavior of some standard estimators of this dimension. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary: 60G55; secondary: 60F05; 60G70; 62F12

Keywords: U-statistic; Poisson random measure; Correlation dimension; Takens estimator; Hill estimator; Stable distribution; Stein–Chen method; Self-normalized sum

1. Introduction

The motivation for considering Poisson convergence of U-statistics in the context of this paper comes from estimators of the so-called correlation dimension which has been studied in chaos theory for some time. Let F denote a distribution on $\mathbb{R}^d$ and let $\|\cdot\|$ denote any norm. Assume that there exists $\alpha > 0$ and a *slowly varying*

* Corresponding author. Fax: +1-6135625776.
E-mail address: ardsg@uottawa.ca (A.R. Dabrowski).

PII: S0304-4149(01)00153-3

(*at infinity*) *function* L (i.e. $L(cx)/L(x) \to 1$ as $x \to \infty$ for any $c > 0$) such that

$$(F \times F) \quad (\{(x_1,x_2): \|x_1 - x_2\| \leqslant x\}) = L(x^{-1})x^{\alpha}. \tag{1.1}$$

Then α is the *correlation dimension of* F, and the probability in (1.1) is called the *correlation integral.*

Given an iid sequence $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots$ of $\mathbb{R}^d$-valued random elements with distribution F, a natural estimator for probability (1.1) is the U-statistic

$$F_n(x) = \binom{n}{2}^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} I_{[0,x]}(\|\mathbf{X}_i - \mathbf{X}_j\|), \tag{1.2}$$

also called the *sample correlation integral*. For practical purposes the assumption of ergodicity for $(\mathbf{X}_i)$ is more appropriate. For x fixed, (1.2) is a consistent estimator of (1.1) even for ergodic sequences $(\mathbf{X}_i)$; see Aaronson et al. (1996). However, the methods of proof used below do not immediately carry over to this more general case.

In practice one often plots $\log F_n(x)$ against $\log x$ for a variety of "small" x-values and estimates α, for example, by a least squares regression procedure; see Grassberger and Procaccia (1983). Clearly, for x "too small" one runs out of sample points and then $F_n(x) = 0$. On the other hand, if x is "too large" one cannot expect that the asymptotic power law behavior of the left tail in (1.1) is guaranteed any more. In other words, one has to deal with a classical semi-parametric problem where the function $L(x)$ is not specified and one has to choose an x-region where one can gain sufficient information about the value α in (1.1) from the sample distribution function F_n. This leads one into the well known semi-parametric problems: biased estimation of α if x is too large and large variance of estimators if x is too small.

In what follows, we consider the above mentioned problems in the context of more general U-statistics. Let $h: \mathbb{R}^{2d} \to \mathbb{R}_+$ be a measurable symmetric function. The positivity restriction on h is not really necessary and is motivated by the above example; it can be relaxed by introducing a probability balancing condition in the neighborhood of zero. Alternatively, the theory below could be formulated in terms of right tails (around ∞) instead of left tails (around zero).

Assume that the relation

$$P(h(\mathbf{X}_1, \mathbf{X}_2) \leqslant x) = L(x^{-1})\, x^{\alpha}, \tag{1.3}$$

for some $\alpha > 0$ and a slowly varying L, holds for some small neighborhood of the origin. The latter regular variation condition is equivalent to the existence of a sequence (a_n) such that $a_n \to \infty$ and

$$P(h(\mathbf{X}_1, \mathbf{X}_2) \leqslant a_n^{-1}) \sim \frac{2}{n^2}, \quad n \to \infty.$$

Clearly, $a_n = n^{2/\alpha}\tilde{L}(n)$ for some slowly varying function $\tilde{L}$; see Bingham et al. (1987, Chapter 1), for more details on regular variation.

It is immediate that the problem of estimating the probability (1.3) for small x is closely related to the lower order statistics in the sample of dependent random variables $h(\mathbf{X}_i, \mathbf{X}_j)$, $1 \leqslant i < j \leqslant n$. A standard approach to the extremes for dependent sequences is via the weak convergence of point processes; see Leadbetter et al. (1983, Chapter 5), Resnick (1987, Section 3.5), or Embrechts et al. (1997, Section 5.2). For this reason we will first study the point processes

$$N_n(\cdot) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \varepsilon_{((i/n, j/n), a_n h(\mathbf{X}_i, \mathbf{X}_j))}(\cdot), \tag{1.4}$$

where $\varepsilon_{(x,y),z}$ denotes the point measure at $((x,y),z)$. We consider the N_n's as random measures on the state space

$$\mathbf{E} = \mathbf{E}_1 \times \mathbb{R}_+ = \{(x,y)\colon 0 < x \leqslant 1, 0 < y < x\} \times \mathbb{R}_+ \tag{1.5}$$

and with values in $M_{\mathrm{p}}(\mathbf{E})$, the space of point measures on $\mathbf{E}$, endowed with the vague topology. We refer to Kallenberg (1983) and Resnick (1987) for the theory of point processes and their weak convergence. Related work, also with similar applications in spatial statistics in mind, is due to Silverman and Brown (1978), Barbour and Eagleson (1984); see also the references therein. The former paper also deals with the problem of estimating the probability (1.1) when $\mathbf{X} \in \mathbb{R}^2$ has a bounded, a.e. continuous density. In this case, it is not difficult to see that the correlation dimension is $\alpha = 2$.

We will show in Section 2 that, under a technical condition, the sequence (N_n) converges to a Poisson process on $\mathbf{E}$ with mean measure (for $0 < a_1 < b_1$, $0 < a_2 < b_2$, $a_3 < b_3$),

$$\eta\left(\bigotimes_{l=1}^{3} (a_l, b_l]\right) = 2(b_1 - a_1)(b_2 - a_2)(b_3^{\alpha} - a_3^{\alpha}). \tag{1.6}$$

This result explains the empirically observed fact that the *dependent* points $a_n h(\mathbf{X}_i, \mathbf{X}_j)$ below a small threshold δ behave very much like an iid sequence of points with the same distribution as $a_n h(\mathbf{X}_i, \mathbf{X}_j)$. Silverman and Brown (1978) showed the same result for squares of interpoint distances $h(\mathbf{X}_i, \mathbf{X}_j) = \|\mathbf{X}_i - \mathbf{X}_j\|^2$ in $\mathbb{R}^2$ when $\mathbf{X}$ has a bounded and a.e. continuous density, thus η is Lebesgue measure and the Poisson process homogeneous on $\mathbf{E}$. In the proof we make heavy use of the elegant Stein–Chen techniques developed by Barbour and Eagleson (1984). Although the generalization to the case of stationary dependent $\mathbf{X}_i$'s (the most interesting case for correlation dimensions) is not straightforward, the results could be extended to U-statistics with kernels on an arbitrary number of variables.

The convergence of point processes has some consequences which are known in the folklore on extremes. We consider some of these in Section 4.

The convergence of the point processes N_n immediately gives one the joint weak limits of the lower order statistics of the $h(\mathbf{X}_i, \mathbf{X}_j)$'s. Under the condition (1.3) it is natural to expect that the Weibull distribution with parameter α determines the limit distribution. This is indeed true and so is analogous to the extreme value theory for stationary sequences; cf. Leadbetter et al. (1983), Resnick (1987) or Embrechts et al. (1997) for details. Joint convergence of the lower extremes is treated in Section 3.1. Similar results were reported earlier in Silverman and Brown (1978). Where condition (1.3) relaxed to require only a distribution in the minimum domain attraction of an extreme value distribution, the same techniques as used for showing the weak convergence of (N_n) could be used for such minimum domains of attraction. This idea is quite straightforward (and standard in extreme value theory) since one can switch from one domain of attraction to another by suitable transformations of the $h(\mathbf{X}_i, \mathbf{X}_j)$'s; see Embrechts et al. (1997, Chapter 3), where such techniques were used for iid sequences.

If $\alpha \in (0, 2)$ one may expect, in analogy with iid sequences with tail index α, that the lower order statistics and the partial sums of the quantities $[a_n h(\mathbf{X}_i, \mathbf{X}_j)]^{-1}$ have the same order of magnitude, and therefore they might have a joint distributional limit; see Resnick (1986) for related work for iid sequences and Davis and Hsing (1995) for sequences of stationary random variables. This idea can be made to work, see Section 3.2.3 where the joint convergence of sums and extremes of the $[a_n h(\mathbf{X}_i, \mathbf{X}_j)]^{-1}$ is considered. Moreover, in Sections 3.2.1 and 3.2.2 we show that the U-statistics with kernel $1/h(\mathbf{x}, \mathbf{y})$ have an α-stable limit provided that $\alpha < 2$. According to our knowledge, these are the first stable limit law results for U-statistics with infinite variance which have been derived from point process convergence. We think that this is a natural approach to the problem of α-stable limits for U-statistics. In contrast to the use of the Hoeffding decomposition, an approach borrowed from L^2 theory (see for example Heinrich and Wolf (1993) and the references therein), our approach exploits the extremal behavior of certain summands, a behaviour which is inherited by the U-statistic.

We started this section with the motivation of estimating the correlation dimension. In Sections 4.1–4.3 we return to this point. Section 4.2 is devoted to the so-called K-function used in spatial statistics for detecting spatial dependencies. In essence, it is the U-statistic (1.2) which, as a function of x, converges in the function space $\mathbf{D}[0, \infty)$. As a consequence, one can derive the limit behavior of this function for very small x-values. A resulting least squares estimator for α based on the K-function for small x is considered as well. In Section 4.1 we consider an estimator of α known in the chaos literature under the name Takens estimator. The latter is a U-statistic based on the quantities $\log h(\mathbf{X}_i, \mathbf{X}_j)$ for very small values $h(\mathbf{X}_i, \mathbf{X}_j)$. We derive the limit distribution of this U-statistic. Finally, in Section 4.3 we consider the Hill estimator which is a standard estimator for α in extreme value theory based on the $\log h(\mathbf{X}_i, \mathbf{X}_j)$'s. We prove the consistency of the estimator based on the asymptotic behavior of the so-called tail empirical process. The latter allows for an increasing (on average) number of small points $h(\mathbf{X}_i, \mathbf{X}_j)$ in any neighborhood of zero and seems to be more appropriate for estimation purposes than the other estimators, under the conditions to be specified in the corresponding sections. This is in contrast to Poisson convergence of the point processes N_n where the average number of points in a neighborhood of zero must be bounded on average.

2. Point process convergence for symmetric functions of iid sequences

In this section we present our main result on the convergence of the point processes N_n defined in (1.4). The points $((i/n, j/n); a_n h(\mathbf{X}_i, \mathbf{X}_j))$ are dependent and linked by the symmetric function h on $\mathbb{R}^{2d}$. Nevertheless, under a mild condition the weak limit of (N_n) turns out to be a point process without multiple points: the Poisson process N with state space $\mathbf{E}$ defined in (1.5), i.e. the limiting points do not cluster in $\mathbf{E}$ and every point occurs at most once with probability one.

The mild condition on the dependence of left tail events related to $h(\mathbf{X}_1, \mathbf{X}_2)$ and $h(\mathbf{X}_1, \mathbf{X}_3)$ is that for any $x > 0$ as $n \to \infty$

$$n^3 P(a_n h(\mathbf{X}_1, \mathbf{X}_2) \leqslant x, a_n h(\mathbf{X}_1, \mathbf{X}_3) \leqslant x) \to 0. \tag{2.1}$$

This condition is not automatic. It fails for X_i independent uniform variables on $[0,1]$, and $h(x_1,x_2) = \max\{x_1,x_2\}$. The condition is the same as used in Barbour and Eagleson (1984) (their condition (2)) and as in Silverman and Brown (1978) (their condition (1)). The former obtained bounds for the total variation distance between the distribution of a U-statistic (with kernel assuming only the values 0 and 1) and an appropriate Poisson distribution.

Condition (2.1) means that in an asymptotic sense very small values of $h(\mathbf{X}_1, \mathbf{X}_2)$ and $h(\mathbf{X}_1, \mathbf{X}_3)$ occur separately from each other, i.e. (2.1) is an anti-clustering condition. This supplements the relation

$$\frac{n^2}{2} P(a_n h(\mathbf{X}_1, \mathbf{X}_2) \leqslant x) \to x^\alpha, \quad x > 0. \tag{2.2}$$

which follows from the regular variation condition (1.3). The following theorem is our main result on point process convergence.

Theorem 2.1. *Let* $(\mathbf{X}_i)$ *be an iid sequence of* $\mathbb{R}^d$*-valued random vectors and* $h : \mathbb{R}^{2d} \to \mathbb{R}_+$ *be a symmetric measurable function. Assume that the regular variation condition* (1.3) *and the asymptotic independence condition* (2.1) *hold. Then* $N_n \xrightarrow{\mathrm{d}} N$, *where* $\xrightarrow{\mathrm{d}}$ *denotes convergence in distribution in* $M_{\mathrm{p}}(\mathbf{E})$ *and* N *is a Poisson process on the state space* $\mathbf{E}$ *with mean measure* η *given by* (1.6).

Proof. Write λ_d for Lebesgue measure on $\mathbb{R}^d$ and $\lambda_{(\alpha)}$ for the measure on $\mathbb{R}_+$ satisfying the relation $\lambda_{(\alpha)}((a,b]) = b^\alpha - a^\alpha$. Since η has a density, the limit process N is simple. Write $\eta_n(\cdot)$ for $EN_n(\cdot)$. Kallenberg's Theorem (see Resnick, 1987, Proposition 3.22 and its remark) states that to prove the theorem it suffices to show the following two conditions:

$$\lim_{n\to\infty} P(N_n(R) = 0) = \mathrm{e}^{-\eta(R)}, \tag{2.3}$$

$$\lim_{n\to\infty} \eta_n(R) = \eta(R), \tag{2.4}$$

where R is any finite union of bounded rectangles

$$\bigcup_{k=1}^{m} A_k \times B_k \subset \mathbf{E}, \quad A_k = (a_k^{(1)}, b_k^{(1)}] \times (a_k^{(2)}, b_k^{(2)}], \quad B_k = (a_k^{(3)}, b_k^{(3)}],$$

where $k = 1, \ldots, m$. It is not difficult to see that one may assume that the A_k's can be chosen disjointly, and we will assume this condition without loss of generality.

We start by checking (2.4). It clearly suffices to consider the case $m = 1$. Since (2.2) holds, we obtain (where $nA = \{(i,j) : (i/n, j/n) \in A\}$)

$$EN_n(A_1 \times B_1) = \sum_{(i,j)\in nA_1} P(a_n h(\mathbf{X}_1, \mathbf{X}_2) \in B_1) \sim 2\lambda_2(A_1)\lambda_{(\alpha)}(B_1) = \eta(R).$$

This proves (2.4).

Now we turn to (2.3). For ease of presentation we will restrict ourselves to the case $m=2$; the general case $m \geqslant 3$ is completely analogous. In fact, at (2.9) below, the computations for $m>2$ reduce to pairwise calculations. Therefore write

$$R=(A_1 \times B_1) \cup (A_2 \times B_2).$$

We will closely follow the arguments given in Barbour and Eagleson (1984); cf. Lee (1990, Section 3.2.4). The argument is based on the Chen–Stein method; see Barbour et al. (1992) for an extensive discussion. We start by observing that we can write

$$N_n(R)=\sum_{i=2}^{n}\sum_{j=1}^{i-1}[I_{nA_1}((i,j))I_{B_1}(a_n h(\mathbf{X}_i,\mathbf{X}_j))+I_{nA_2}((i,j))I_{B_2}(a_n h(\mathbf{X}_i,\mathbf{X}_j))].$$

Write P_n for the Poisson distribution on $\mathbf{N}=\{0,1,2,\ldots\}$ with mean $\eta_n=\eta_n(R)$ (suppressing the dependence on R). According to (2.4),

$$\begin{aligned}
&|P(N_n(R)=0)-P(N(R)=0)|\\
&\quad\leqslant |P(N_n(R)=0)-P_n(0)|+|P_n(0)-P(N(R)=0)|\\
&\quad=|P(N_n(R)=0)-P_n(0)|+\mathrm{o}(1).
\end{aligned}$$

Define a real-valued function x (suppressing the dependence on n) on the non-negative integers as follows:

$$\begin{aligned}
&x(0)=0,\\
&x(m+1)=\mathrm{e}^{\eta_n}\frac{m!}{\eta_n^{m+1}}[P_n(\{0\})-P_n(\{0\})P_n(\{0,\ldots,m\})], \quad m=0,1,\ldots\ .
\end{aligned}$$

The function x has the following properties (see Barbour and Eagleson, 1984, p. 400): x is bounded and satisfies

$$\Delta x=\sup_m |x(m+1)-x(m)|<\min(1,(\eta_n)^{-1}). \tag{2.5}$$

It follows from Barbour and Eagleson (1984, p. 401), that

$$|P(N_n(R)=0)-P_n(0)|\leqslant |E[\eta_n x(N_n(R)+1)-N_n(R)x(N_n(R))]|. \tag{2.6}$$

Write

$$\begin{aligned}
&\mathbf{D}=\{\mathbf{k}:\ \boldsymbol{k}=(k_1,k_2), 1\leqslant k_1<k_2\leqslant n\},\\
&I_{\mathbf{k}}=I_{nA_1}(\mathbf{k})I_{B_1}(a_n h(\mathbf{X}_{k_1},\mathbf{X}_{k_2}))+I_{nA_2}(\mathbf{k})I_{B_2}(a_n h(\mathbf{X}_{k_1},\mathbf{X}_{k_2})),\\
&\eta_{\mathbf{k}}=EI_{\mathbf{k}},
\end{aligned}$$

where we suppress in the notation the dependence on n. For a fixed $\mathbf{k}\in\mathbf{D}$ we decompose $\mathbf{D}$ into $\mathbf{D}_{1\mathbf{k}}=\{\ell\in\mathbf{D}:\ \ell_i\neq k_j,\ i,j=1,2\}$ and $\mathbf{D}_{2\mathbf{k}}=\{\ell\in\mathbf{D}:\ \ell\neq\mathbf{k},\ \ell_i=k_j$

for some $i,j=1,2\}$. Then

$$
\begin{aligned}
N_n(R) &= \sum_{\ell\in\mathbf{D}} I_\ell \\
&=: \sum_{\ell\in\mathbf{D}_{1\mathbf{k}}} I_\ell + \left(I_{\mathbf{k}} + \sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} I_\ell \right) \\
&=: N_n^{(1)}(\mathbf{k}) + N_n^{(2)}(\mathbf{k}).
\end{aligned}
$$

We may expand the right-hand side of (2.6) to obtain

$$
\begin{aligned}
|P(N_n(R)=0)-P_n(0)| \leqslant{} & \left| \sum_{\ell\in\mathbf{D}} \eta_\ell E[x(N_n(R)+1) - x(N_n^{(1)}(\ell)+1)] \right| \\
& + \left| \sum_{\ell\in\mathbf{D}} [E(I_\ell x(N_n(R)) - \eta_\ell E(x(N_n^{(1)}(\ell)+1))] \right| \\
=:{} & J_1 + J_2. \qquad (2.7)
\end{aligned}
$$

We now follow the argument on p. 401 of Barbour and Eagleson (1984). From property (2.5) of the x-function, we conclude that

$$
\begin{aligned}
J_1 &\leqslant \sum_{\mathbf{k}\in\mathbf{D}} \eta_{\mathbf{k}} \sum_{k=0}^{\infty} |E(x(N_n(R)+1) - x(N_n^{(1)}(\mathbf{k})+1) \mid N_n^{(2)}(\mathbf{k})=k)| P(N_n^{(2)}(\mathbf{k})=k) \\
&= \sum_{\mathbf{k}\in\mathbf{D}} \eta_{\mathbf{k}} \sum_{k=0}^{\infty} |E(x(N_n^{(1)}(\mathbf{k}) + N_n^{(2)}(\mathbf{k})+1) \\
&\qquad - x(N_n^{(1)}(\mathbf{k})+1) \mid N_n^{(2)}(\mathbf{k})=k)| P(N_n^{(2)}(\mathbf{k})=k) \\
&= \sum_{\mathbf{k}\in\mathbf{D}} \eta_{\mathbf{k}} \sum_{k=0}^{\infty} |E(x(N_n^{(1)}(\mathbf{k}) + k+1) \\
&\qquad - x(N_n^{(1)}(\mathbf{k})+1) \mid N_n^{(2)}(\mathbf{k})=k)| P(N_n^{(2)}(\mathbf{k})=k) \\
&\leqslant \Delta x \sum_{\mathbf{k}\in\mathbf{D}} \eta_{\mathbf{k}} \sum_{k=0}^{\infty} k P(N_n^{(2)}(\mathbf{k})=k) = \Delta x \sum_{\mathbf{k}\in\mathbf{D}} \eta_{\mathbf{k}} E N_n^{(2)}(\mathbf{k}). \qquad (2.8)
\end{aligned}
$$

Writing

$$
m_n(A_j) = \frac{1}{n^2} \sum_{\mathbf{k}\in\mathbf{D}} I_{nA_j}(\mathbf{k}), \quad j=1,2,
$$

and using the definitions of I_ℓ and $\eta_\ell = EI_\ell$, we obtain that the last sum in (2.8) equals

$$\sum_{\mathbf{k}\in\mathbf{D}}\left(\sum_{i=1}^{2} I_{nA_i}(\mathbf{k})P(a_n h(\mathbf{X}_{k_1},\mathbf{X}_{k_2})\in B_i)\right)\left(\sum_{\ell\in\mathbf{D}}\sum_{j=1}^{2} I_{nA_j}(\ell)P(a_n h(\mathbf{X}_{\ell_1},\mathbf{X}_{\ell_2})\in B_j)\right.$$
$$\left.-\sum_{\ell\in\mathbf{D}_{1\mathbf{k}}}\sum_{j=1}^{2} I_{nA_j}(\ell)P(a_n h(\mathbf{X}_{\ell_1},\mathbf{X}_{\ell_2})\in B_j)\right)$$
$$=\sum_{i=1}^{2}\sum_{j=1}^{2}[n^2P(a_nh(\mathbf{X}_1,\mathbf{X}_2)\in B_i)][n^2P(a_nh(\mathbf{X}_1,\mathbf{X}_2)\in B_j)]m_n(A_i)m_n(A_j)$$
$$\times\left(1-[m_n(A_i)m_n(A_j)]^{-1}\frac{1}{n^4}\sum_{\mathbf{k}\in\mathbf{D}}\sum_{\ell\in\mathbf{D}_{1\mathbf{k}}} I_{nA_i}(\mathbf{k})I_{nA_j}(\ell)\right). \tag{2.9}$$

By the regular variation condition (1.3),

$$n^2P(a_nh(\mathbf{X}_1,\mathbf{X}_2)\in B_i)\to 2\lambda_{(\alpha)}(B_i),\quad i=1,2.$$

Moreover since ℓ and $\mathbf{k}$ are groups of indices,

$$m_n(A_i)\to\lambda_2(A_i),\quad i=1,2,$$
$$\frac{1}{n^4}\sum_{\mathbf{k}\in\mathbf{D}}\sum_{\ell\in\mathbf{D}_{1\mathbf{k}}} I_{nA_i}(\mathbf{k})I_{nA_j}(\ell)\to\lambda_2(A_i)\lambda_2(A_j),\quad i,j=1,2.$$

Therefore the right-hand side in (2.9) converges to zero as $n\to\infty$, and so we conclude that $J_1\to 0$.

By (2.7) it remains to prove $J_2\to 0$. Since $N_n^{(1)}(\ell)$ and I_ℓ are independent, we have

$$J_2=\left|\sum_{\mathbf{k}\in\mathbf{D}} E[I_{\mathbf{k}}x(N_n(R))-E(I_{\mathbf{k}}x(N_n^{(1)}(\mathbf{k})+1))]\right|$$
$$=\left|\sum_{\mathbf{k}\in\mathbf{D}} E[I_{\mathbf{k}}(x(N_n(R))-x(N_n^{(1)}(\mathbf{k})+1))]\right|.$$

A conditioning argument (but $I_{\mathbf{k}}$ and $N_n^{(2)}(\mathbf{k})$ are not independent) similar to (2.8) gives

$$J_2\leqslant\Delta x\sum_{\mathbf{k}\in\mathbf{D}} E[I_{\mathbf{k}}(N_n^{(2)}(\mathbf{k})-1)]\leqslant\Delta x\sum_{\mathbf{k}\in\mathbf{D}} E\left[I_{\mathbf{k}}\sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} I_\ell\right]$$
$$=\Delta x\sum_{\mathbf{k}\in\mathbf{D}}\sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} E\left[\left(\sum_{i=1}^{2} I_{nA_i}(\mathbf{k})I_{B_i}(a_nh(\mathbf{X}_{k_1},\mathbf{X}_{k_2}))\right)\right.$$
$$\left.\times\left(\sum_{j=1}^{2} I_{nA_j}(\ell)I_{B_j}(a_nh(\mathbf{X}_{\ell_1},\mathbf{X}_{\ell_2}))\right)\right]$$

$$= \Delta x \sum_{\mathbf{k}\in\mathbf{D}} \sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} \sum_{i=1}^{2} \sum_{j=1}^{2} I_{nA_i}(\mathbf{k}) I_{nA_j}(\ell) P(a_n h(\mathbf{X}_1,\mathbf{X}_2)\in B_i, a_n h(\mathbf{X}_1,\mathbf{X}_3)\in B_j)$$

$$= \Delta x \sum_{i=1}^{2} \sum_{j=1}^{2} \left[\frac{1}{n^3} \sum_{\mathbf{k}\in\mathbf{D}} \sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} I_{nA_i}(\mathbf{k}) I_{nA_j}(\ell) \right.$$

$$\left. \times [n^3 P(a_n h(\mathbf{X}_1,\mathbf{X}_2)\in B_i, a_n h(\mathbf{X}_1,\mathbf{X}_3)\in B_j)] \right]. \tag{2.10}$$

By virtue of (2.1) and since

$$\frac{1}{n^3} \sum_{\mathbf{k}\in\mathbf{D}} \sum_{\ell\in\mathbf{D}_{2\mathbf{k}}} I_{nA_i}(\mathbf{k}) I_{nA_j}(\ell) = \mathrm{O}(1),$$

we conclude that the right-hand side in (2.10) converges to zero. This finishes the proof. □

Our first corollary of Theorem 2.1 follows from Lemma 9.1.X in Daley and Vere-Jones (1988). It states that weak convergence of the point processes ξ_n in $M_{\mathrm{p}}((0,\infty))$ implies weak convergence of the cumulative processes $(\xi_n((0,x]))_{x>0}$ in $\mathbf{D}(0,\infty)$ equipped with the J_1-topology. Under the conditions of Theorem 2.1 this implies the following.

Corollary 2.2.

$$N_n(\mathbf{E}_1 \times (0,\delta]) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} I_{(0,\delta]}(a_n h(\mathbf{X}_i,\mathbf{X}_j)) \xrightarrow{\mathrm{d}} N(\mathbf{E}_1 \times (0,\delta]), \quad \delta > 0, \tag{2.11}$$

in $\mathbf{D}(0,\infty)$ *where the right hand process is Poisson on* $(0,\infty)$ *with intensity* $\alpha x^{\alpha-1}\,\mathrm{d}x$.

In particular, Corollary 2.2 establishes the convergence of the tail of U-processes to a Poisson limit. The same approach gives the following result.

Corollary 2.3. *Let A be a bounded Borel set in* $\mathbb{R}_+$. *Then*

$$N_n(\{(0,t] \times (0,1] \times A\} \cap \mathbf{E})$$

$$= \sum_{i=2}^{[nt]} \sum_{j=1}^{i-1} I_A(a_n h(\mathbf{X}_i,\mathbf{X}_j)) \xrightarrow{\mathrm{d}} N(\{(0,t] \times (0,1] \times A\} \cap \mathbf{E}),$$

in $\mathbf{D}(0,1]$ *where the right hand process is a Poisson process on* $(0,1]$ *with mean measure of* $(0,t]$ *equal to* $t^2\lambda_{(\alpha)}(A)$.

Example 2.4. If we want to apply the results of this section to the sample correlation integral (1.2), we need to verify (2.1). Observe that by Fubini's theorem, we may write the correlation integral as

$$\int_{\mathbb{R}^d} F(B_x(\mathbf{y}))\,\mathrm{d}F(\mathbf{y}),$$

where $B_x(\mathbf{y}) = \{\mathbf{z} \in \mathbb{R}^d\colon \|\mathbf{z}-\mathbf{y}\| \leqslant x\}$ is the ball with radius x and center $\mathbf{y}$.

Let (a_n) be chosen that

$$\int_{\mathbb{R}^d} F(B_{1/a_n}(\mathbf{y}))\,\mathrm{d}F(y) \sim 2n^{-2}.$$

Again applying Fubini's theorem, we see that (2.1) reads

$$\int_{\mathbb{R}^d} [F(B_{1/a_n}(\mathbf{y})]^2\,\mathrm{d}F(\mathbf{y}) = \mathrm{o}(n^{-3}).$$

This will hold, for example, if F has a bounded density with respect to Lebesgue measure.

3. Weak convergence

In this section we use the point process convergence result to establish weak convergence results for the extreme values of $h(\mathbf{X}_i, \mathbf{X}_j)$ and stable limits for sums for such terms.

3.1. Joint convergence of the extremes

Consider the order statistics

$$h_{(1)} \leqslant \cdots \leqslant h_{(n(n-1)/2)}$$

of the $n(n-1)/2$ values $(h(\mathbf{X}_i, \mathbf{X}_j))_{1 \leqslant i < j \leqslant n}$. It follows from the folklore in extreme value theory that convergence of the point processes N_n implies the joint weak convergence of the vectors $a_n(h_{(1)}, \ldots, h_{(k)})$. Indeed, for $0 < x_k < \cdots < x_1$,

$$\begin{aligned}
&P((a_n h_{(1)})^{-1} \leqslant x_1, (a_n h_{(2)})^{-1} \leqslant x_2, \ldots, (a_n h_{(k)})^{-1} \leqslant x_k)\\
&\quad = P(N_n(\mathbf{E}_1 \times [0, x_1^{-1}]) = 0, N_n(\mathbf{E}_1 \times [0, x_2^{-1}]) \leqslant 1, \ldots, N_n(\mathbf{E}_1 \times [0, x_k^{-1}])\\
&\qquad \leqslant k-1)\\
&\quad \to P(N(\mathbf{E}_1 \times [0, x_1^{-1}]) = 0, N(\mathbf{E}_1 \times [0, x_2^{-1}]) \leqslant 1, \ldots, N(\mathbf{E}_1 \times [0, x_k^{-1}])\\
&\qquad \leqslant k-1).
\end{aligned}$$

The limiting density φ_α of the vector $[a_n(h_{(1)}, \ldots, h_{(k)})]^{-1}$ can now be read off from Example 4.2.9 of Embrechts et al. (1997):

$$\varphi_\alpha(x_1, \ldots, x_k) = \alpha^k \exp\left\{ -x_k^{-\alpha} - (\alpha+1) \sum_{j=1}^{k} \log x_j \right\}, \quad 0 < x_k < \cdots < x_1.$$

The distribution of the lower order statistics $a_n(h_{(1)}, \ldots, h_{(k)})$ can be obtained by a straightforward transformation of the density φ_α.

3.2. Stable laws for heavy-tailed kernels

Weak convergence results for the U-statistics

$$H_n = \sum_{i=2}^{n}\sum_{j=1}^{i-1} h(\mathbf{X}_i, \mathbf{X}_j),$$

when $\mathrm{var}(h(\mathbf{X}_1, \mathbf{X}_2)) < \infty$ are well-known; see for example Serfling (1980). To the authors' knowledge, no stable limits using point process results have been established in the case of infinite variances. Stable limits for sums of independent (or weakly dependent) random variables have a long history, but some recent work has focused on an approach via point processes (see Resnick, 1986; Davis and Hsing, 1995). Once a result like Theorem 2.1 is obtained, applications of the continuous mapping theorem and separate arguments for the cases where the stable index is less than 1, equal to 1, and between 1 and 2 will yield the weak convergence. Here we establish the stable limits for the first and third cases; the case where $\alpha = 1$ remains open.

Whereas we considered left tails of positive kernels h earlier, here we need to consider heavy right-tailed kernels, g. Obviously, one can translate from one to the other by $g = 1/h$, but we also need the corresponding weak convergence of the point processes.

We may conclude from Theorem 2.1 that in $M_{\mathrm{p}}(\mathbf{E})$

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1} \varepsilon_{((i/n, j/n), a_n h(\mathbf{X}_i, \mathbf{X}_j))}(\cdot) \xrightarrow{\mathrm{d}} N(\cdot) \overset{\mathrm{d}}{=} \sum_{k=1}^{\infty} \varepsilon_{(\mathbf{U}_k, \Gamma_k^{1/\alpha})}(\cdot), \tag{3.1}$$

where $\varepsilon_{(x,y),z}$ denotes a point measure at $((x,y),z) \in \mathbf{E}$, N is a Poisson process on $\mathbf{E}$ with mean measure $\eta = 2\lambda_2 \times \lambda_{(\alpha)}$ which has representation through the points $\mathbf{U}_k$ of an iid sequence of uniform (on $\mathbf{E}_1$) random vectors and the points Γ_k of a unit rate Poisson process on $\mathbb{R}_+$, independent of $(\mathbf{U}_k)$. The values of a_n and α are determined from h as before.

For convenience we write

$$g_n(\mathbf{X}_i, \mathbf{X}_j) = a_n^{-1} g(\mathbf{X}_i, \mathbf{X}_j) = [a_n h(\mathbf{X}_i, \mathbf{X}_j)]^{-1}.$$

It is an immediate consequence of the continuous mapping theorem that the point processes of large values of g converge weakly, i.e.

$$\tilde{N}_n(\cdot) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} \varepsilon_{((i/n, j/n), g_n(\mathbf{X}_i,\ \mathbf{X}_j))}(\mathbf{E}_1 \times \cdot) \xrightarrow{\mathrm{d}} \tilde{N}(\cdot) = \sum_{k=1}^{\infty} \varepsilon_{\Gamma_k^{-1/\alpha}}(\cdot),$$

where the convergence holds in $M_{\mathrm{p}}((0,\infty])$ and $\tilde{N}$ is a Poisson process with mean measure of $(x,\infty]$ equal to $x^{-\alpha}$. Now consider the functional $T_\delta : M_{\mathrm{p}}((0,\infty]) \to \mathbb{R}$ for a fixed $\delta > 0$ given by

$$T_\delta\left(\sum_{k=1}^{\infty} \varepsilon_{z_k}\right) = \sum_{k=1}^{\infty} z_k I_{(\delta,\infty]}(z_k).$$

This is an almost surely (with respect to N) continuous functional; see Resnick (1986). Thus we obtain by the continuous mapping theorem that for every $\delta > 0$,

$$\tilde{G}_n(\delta) := \sum_{i=2}^{n}\sum_{j=1}^{i-1} g_n(\mathbf{X}_i, \mathbf{X}_j) I_{(\delta,\infty]}(g_n(\mathbf{X}_i, \mathbf{X}_j))$$

$$\xrightarrow{\mathrm{d}} \tilde{G}(\delta) = \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} I_{(\delta,\infty]}(\Gamma_k^{-1/\alpha}). \tag{3.2}$$

This line holds for every $\alpha > 0$. If $\alpha \in (0,1)$ the right-hand series (for $\delta = 0$) characterizes an α-stable law on $\mathbb{R}_+$; see Samorodnitsky and Taqqu (1994, Sections 1.4 and 1.5). To let $\delta \downarrow 0$, i.e. to prove a weak convergence result for the U-statistics

$$G_n = \sum_{i=2}^{n}\sum_{j=1}^{i-1} g_n(\mathbf{X}_i, \mathbf{X}_j),$$

one has to treat the cases $\alpha \in (0,1)$, $\alpha = 1$, $\alpha \in (1,2)$ and $\alpha > 2$ in different ways. In the last case, $\operatorname{var}(g(\mathbf{X}_1, \mathbf{X}_2)) < \infty$ and therefore the standard limit theory for U-statistics is available; see for example Serfling (1980). The case $\alpha = 1$ is traditionally the most sensitive, and we do not address it here.

3.2.1. *The case $\alpha \in (0,1)$*

Theorem 3.1. *For $\alpha \in (0,1)$, the following relation holds*:

$$G_n \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha}.$$

The limiting infinite series converges almost surely to a positive α-stable random variable.

Proof. The existence of the almost sure limit of $\tilde{G}(\delta)$ as $\delta \downarrow 0$ and the fact that it represents an α-stable random variable is well-known; cf. Samorodnitsky and Taqqu (1994, Sections 1.4 and 1.5). Write

$$G_n(\delta) = G_n - \tilde{G}_n(\delta).$$

It remains to show (see Billingsley, 1968, Theorem 4.2) that

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} P(G_n(\delta) > \gamma) = 0 \quad \text{for all } \gamma > 0. \tag{3.3}$$

Since $g(\mathbf{X}_1, \mathbf{X}_2) = (h(\mathbf{X}_1, \mathbf{X}_2))^{-1}$ is a random variable with regularly varying tail of index $-\alpha$, we have by Karamata's theorem (cf. Bingham et al., 1987, Section 1.6) that as $n \to \infty$,

$$EG_n(\delta) = \frac{n(n-1)}{2a_n} E[g(\mathbf{X}_1, \mathbf{X}_2) I_{\{g(\mathbf{X}_1,\mathbf{X}_2) \leqslant a_n\delta\}}] \sim \frac{\alpha}{1-\alpha}\delta^{1-\alpha},$$

which, together with Markov's inequality, implies (3.3). □

3.2.2. The case $\alpha \in (1,2)$

Note that $g(\mathbf{X}_1, \mathbf{X}_2)$ is a random variable with regularly varying tail of index α. Thus, if $\alpha > 1$, this random variable has a finite first moment, and so it becomes necessary to center the U-statistics $G_n(\delta)$ and G_n by their means. Moreover, it is not difficult to check that the limit of $\tilde{G}(\delta)$ as $\delta \downarrow 0$ does not exist any longer unless one centers $\tilde{G}(\delta)$ as well.

Write for $\delta > 0$,

$$g_n^{(\delta)}(\mathbf{X}_i, \mathbf{X}_j) := g_n(\mathbf{X}_i, \mathbf{X}_j) I_{[0,\delta]}(g_n(\mathbf{X}_i, \mathbf{X}_j)).$$

Theorem 3.2. *Assume* $\alpha \in (1,2)$ *and*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} n^3 \operatorname{cov}(g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_2), g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_3)) = 0. \tag{3.4}$$

Then

$$G_n - EG_n \xrightarrow{\mathrm{d}} \lim_{\delta \downarrow 0} (G(\delta) - EG(\delta)),$$

where $G(\delta)$ *is defined in* (3.2). *The right-hand limit exists almost surely and represents an* α*-stable random variable which is totally skewed to the right.*

We start with an auxiliary result about the variance of $G_n(\delta)$.

Lemma 3.3. *Assume that* (3.4) *holds. Then*

$$\operatorname{var}(G_n(\delta)) \sim n^3 \operatorname{cov}(g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_2), g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_3)) + \frac{\alpha}{2-\alpha} \delta^{2-\alpha}.$$

Proof. Since $G_n(\delta)$ is a U-statistic we can apply standard results for the variance of $G_n(\delta)$; cf. Lee (1990, p. 12)

$$\operatorname{var}(G_n(\delta)) \sim n^3 \operatorname{cov}(g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_2), g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_3)) + \frac{n^2}{2} \operatorname{var}(g_n^{(\delta)}(\mathbf{X}_1, \mathbf{X}_2)). \tag{3.5}$$

By Karamata's theorem, the second term on the right-hand side is asymptotically of the order

$$\frac{\alpha}{2-\alpha} \frac{n^2}{2a_n^2} (\delta a_n)^2 P(g(\mathbf{X}_1, \mathbf{X}_2) a_n^{-1} > \delta) \sim \frac{\alpha}{2-\alpha} \delta^{2-\alpha}. \qquad \square$$

Proof of Theorem 3.2. We proceed in a way similar to the proof of Theorem 3.1. Notice that by regular variation,

$$E\tilde{G}_n(\delta) = \frac{n(n-1)}{2a_n} E(g(\mathbf{X}_1, \mathbf{X}_2) I_{(\delta,\infty)}(g_n(\mathbf{X}_1, \mathbf{X}_2))) \sim \delta^{1-\alpha} \frac{\alpha}{\alpha - 1}.$$

From the latter relation and (3.2) we obtain that for every $\delta > 0$,

$$\tilde{G}_n(\delta) - E\tilde{G}_n(\delta) \xrightarrow{\mathrm{d}} \tilde{G}(\delta) - \delta^{1-\alpha} \frac{\alpha}{\alpha - 1} = \tilde{G}(\delta) - E\tilde{G}(\delta).$$

The right-hand side, however, converges as $\delta \downarrow 0$ to an α-stable random variable which is totally skewed to the right; see Samorodnitsky and Taqqu (1994, Sections 1.4 and 1.5). Thus it remains to show that

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P(|G_n(\delta) - EG_n(\delta)| > \gamma) = 0 \quad \text{for all } \gamma > 0.$$

But this follows from Lemma 3.3 and an application of Chebyshev's inequality. □

To apply Theorem 3.2 one needs to verify condition (3.4). The situation is much simpler if one considers U-statistics based on the random variables

$$\hat{g}(\mathbf{X}_i, \mathbf{X}_j) = r_{ij} g(\mathbf{X}_i, \mathbf{X}_j), \quad i = 2, 3, \ldots, j = 1, \ldots, i-1,$$

where (r_{ij}) is a sequence of iid Rademacher random variables, i.e. $P(r_{ij} = \pm 1) = 0.5$, independent of $(\mathbf{X}_i)$. Write

$$\hat{g}_n(\mathbf{X}_i, \mathbf{X}_j) = a_n^{-1}\hat{g}(\mathbf{X}_i, \mathbf{X}_j).$$

Theorem 3.4. *If* $\alpha \in (0, 2)$, *then*

$$a_n^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\hat{g}(\mathbf{X}_i, \mathbf{X}_j) \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} r_k \Gamma_k^{-1/\alpha}.$$

The right-hand limit exists almost surely and represents a symmetric α*-stable random variable.*

Proof. In this case, the variance of

$$\hat{G}_n(\delta) = \sum_{i=2}^{n}\sum_{j=1}^{i-1}\hat{g}_n(\mathbf{X}_i, \mathbf{X}_j) I_{[0,\delta]}(\hat{g}_n(\mathbf{X}_i, \mathbf{X}_j))$$

is given by

$$\frac{n(n-1)}{2}\operatorname{var}(\hat{g}_n(\mathbf{X}_1, \mathbf{X}_2) I_{[0,\delta]}(\hat{g}_n(\mathbf{X}_1, \mathbf{X}_2))) \sim \frac{\alpha}{2-\alpha}\delta^{2-\alpha}. \tag{3.6}$$

Following the lines of the proofs above, one can show the weak convergence of the point processes

$$\hat{N}_n(\cdot) = \sum_{i=1}^{n}\sum_{j=1}^{i-1}\varepsilon_{(i/n, j/n), a_n r_{ij} h(\mathbf{X}_i, \mathbf{X}_j)}(\cdot) \xrightarrow{\mathrm{d}} \hat{N}(\cdot)$$

in $M_{\mathrm{p}}(\hat{\mathbf{E}})$, where $\hat{\mathbf{E}} = \mathbf{E}_1 \times \mathbb{R}$ and $\hat{N}$ is a Poisson process on $\hat{\mathbf{E}}$ with mean measure $\hat{\eta}$ given by

$$\mathrm{d}\hat{\eta}((x, y), z) = 0.5\alpha\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z[I_{[0,\infty)}(z) z^{\alpha-1} + I_{(-\infty,0]}(z)(-z)^{\alpha-1}].$$

Then the same arguments leading to (3.2) prove that for every $\alpha > 0$ and $\delta > 0$,

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\hat{g}_n(\mathbf{X}_i, \mathbf{X}_j) I_{(\delta,\infty]}(|\hat{g}_n(\mathbf{X}_i, \mathbf{X}_j)|) \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} r_k \Gamma_k^{-1/\alpha} I_{(\delta,\infty]}(\Gamma_k^{-1/\alpha}). \tag{3.7}$$

Moreover, (3.6) and the symmetry of the random variables $\hat{g}(\mathbf{X}_i, \mathbf{X}_j)$ show that one can let $\delta \downarrow 0$ in (3.7), both on the left- and right-hand sides. Hence

$$a_n^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \hat{g}(\mathbf{X}_i, \mathbf{X}_j) \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} r_k \Gamma_k^{-1/\alpha}.$$

Moreover, the right-hand side represents a symmetric α-stable random variable; see Samorodnitsky and Taqqu (1994, Sections 1.4 and 1.5). □

3.2.3. Self-normalized sums

In the above limit theorems one has to know the value α in order to apply those theorems. In practice it is unknown, but an adaptive random normalization can yet be used to obtain a limit distribution. We used two almost surely continuous mappings both acting on one and the same point process and so one can consider the two mappings as one almost surely continuous mapping; see Resnick (1986, Section 4, in particular Sections 4.6 and 4.7), where the methodology of this approach is explained in detail. For example, in the case $\alpha < 1$ we have

$$\left(\max_{i,j=1,\dots,n} a_n^{-1} g(\mathbf{X}_i, \mathbf{X}_j),\ a_n^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} g(\mathbf{X}_i, \mathbf{X}_j) \right)$$

$$\xrightarrow{\mathrm{d}} \left(\max_{i \geqslant 1} \Gamma_i^{-1/\alpha}, \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha} \right) = \left(\Gamma_1^{-1/\alpha}, \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha} \right).$$

This and a similar argument in the case $\alpha \in (1,2)$ establishes

Theorem 3.5. *If* $0 < \alpha < 1$,

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{g(\mathbf{X}_i, \mathbf{X}_j)}{\max_{l,k=1,\dots,n} g(\mathbf{X}_l, \mathbf{X}_k)} \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} (\Gamma_1/\Gamma_k)^{1/\alpha}.$$

If $1 < \alpha < 2$, *under the conditions of Theorem* 3.2,

$$\left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} g(\mathbf{X}_i, \mathbf{X}_j) - \frac{n(n-1)}{2} Eg(\mathbf{X}_1, \mathbf{X}_2) \right) \Big/ \max_{i,j=1,\dots,n} g(\mathbf{X}_i, \mathbf{X}_j)$$

$$\xrightarrow{\mathrm{d}} \Gamma_1^{1/\alpha} \lim_{\delta \uparrow \infty} \left(\sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} I_{[0,\delta]}(\Gamma_k) - E\left[\sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} I_{[0,\delta]}(\Gamma_k) \right] \right).$$

4. Applications

In this section we consider some applications of the point process convergence result of Theorem 2.1. In what follows, we assume that the assumptions of this theorem are satisfied. We also use the notation of the previous sections, in particular from (3.1).

4.1. Takens's estimator for the correlation dimension

As an alternative to the Grassberger–Procaccia estimator, Takens (1985) introduced a dimension estimator motivated by the maximum likelihood principle. To derive the Takens estimator, assume for a moment that in (1.1) exact scaling holds in some neighborhood of zero, i.e.

$$P(\|\mathbf{X}_1-\mathbf{X}_2\|\leqslant x)=x^{\alpha}\quad \text{for } 0\leqslant x\leqslant\delta.$$

If $R_1,\dots,R_N$ denote independent copies of the distance $\|\mathbf{X}_1-\mathbf{X}_2\|$, we obtain

$$P(R_i\leqslant x\,|\,R_i\leqslant\delta)=(x/\delta)^{\alpha},$$

i.e. conditionally upon $\{R_i\leqslant\delta\}$, the density of R_i is $f(x)=\alpha(x/\delta)^{\alpha-1}$. This fact allows one to derive the maximum likelihood estimator $\hat\alpha$ of α, based on the distances $R_i\leqslant\delta$:

$$\hat\alpha=\left[-\frac{1}{\#\{i\leqslant N: R_i\leqslant\delta\}}\sum_{i=1}^{N}\log(R_i/\delta)I_{[0,\delta]}(R_i)\right]^{-1}.$$

Substituting observed inter-point distances $\|\mathbf{X}_i-\mathbf{X}_j\|$ for the R_i's, one obtains *Takens' estimator*

$$\hat\alpha_{\mathrm{T}}=\left[\frac{-\sum_{i=2}^{n}\sum_{j=1}^{i-1}\log(\|\mathbf{X}_i-\mathbf{X}_j\|/\delta)I_{[0,\delta]}(\|\mathbf{X}_i-\mathbf{X}_j\|)}{\sum_{i=2}^{n}\sum_{j=1}^{i-1}I_{[0,\delta]}(\|\mathbf{X}_i-\mathbf{X}_j\|)}\right]^{-1}.$$

Note that the inter-point distances $\|\mathbf{X}_i-\mathbf{X}_j\|$ are no longer independent so that $\hat\alpha_{\mathrm{T}}$ is not a maximum likelihood estimator. In addition, the Takens estimator is applied in situations where only asymptotic scaling in the sense of (1.1) holds.

The derivation of the Takens estimator was carried out for a fixed sample size n. As n increases, it seems reasonable to use shrinking neighborhoods $[0,\delta_n]$ rather than fixed intervals $[0,\delta]$. In the spirit of our previous results we choose $\delta_n=\delta/a_n$ and are thus led to consider

$$\frac{-\sum_{i=2}^{n}\sum_{j=1}^{i-1}\log(a_nh(\mathbf{X}_i,\mathbf{X}_j)/\delta)I_{[0,\delta]}(a_nh(\mathbf{X}_i,\mathbf{X}_j))}{\sum_{i=2}^{n}\sum_{j=1}^{i-1}I_{[0,\delta]}(a_nh(\mathbf{X}_i,\mathbf{X}_j))}.\tag{4.1}$$

Since

$$P(a_n\min_{1\leqslant i<j\leqslant n}h(\mathbf{X}_i,\mathbf{X}_j)\leqslant\delta)\to P(\Gamma_1\leqslant\delta^{\alpha})>0,$$

it is not unlikely that ratio (4.1) is not well defined. If the denominator in (4.1) is zero, we interpret quantity (4.1) as zero.

As in the previous section, we may conclude by a continuous mapping argument that the above statistic converges in distribution to

$$\frac{-1/\alpha\sum_{k=1}^{\infty}\log(\Gamma_k)I_{[0,\delta^{\alpha}]}(\Gamma_k)}{\sum_{k=1}^{\infty}I_{[0,\delta^{\alpha}]}(\Gamma_k)}+\log\delta.\tag{4.2}$$

As for (4.1), we interpret (4.2) as zero if the denominator is zero. Write N for the unit rate Poisson process generated by the Γ_k's. Then the limiting ratio on the right-hand side can be written as

$$\frac{1}{\alpha}\frac{-\sum_{k=1}^{N(\delta^\alpha)}\log(\Gamma_k)}{N(\delta^\alpha)} =: \frac{S_{N(\delta^\alpha)}}{N(\delta^\alpha)}.$$

The joint distribution of $(S_{N(\delta^\alpha)}, N(\delta^\alpha))$ can be calculated as follows: for $x>0$ and $k \geqslant 1$,

$$P(S_{N(\delta^\alpha)} \leqslant x, N(\delta^\alpha)=k) = P(S_{N(\delta^\alpha)} \leqslant x \mid N(\delta^\alpha)=k)P(N(\delta^\alpha)=k).$$

Given $N(\delta^\alpha)=k$, $(\Gamma_i)_{i=1,\dots,k}$ has the same distribution as the order statistics of a uniform on $[0,\delta^\alpha]$ iid sample $\delta^\alpha(U_i)_{i=1,\dots,k}$. Notice that $-\log(U_i)$ has the standard exponential distribution, and therefore for an independent copy (Γ_k') of (Γ_k),

$$\begin{aligned}P(S_{N(\delta^\alpha)} \leqslant x \mid N(\delta^\alpha)=k) &= P\left(-k\log(\delta^\alpha) - \sum_{i=1}^{k}\log U_i \leqslant x \,\middle|\, N(\delta^\alpha)=k\right)\\ &= P(-k\log(\delta^\alpha) + \Gamma_k' \leqslant x \mid N(\delta^\alpha)=k)\\ &= P(-k\log(\delta^\alpha) + \Gamma_k' \leqslant x).\end{aligned}$$

Hence for $k \geqslant 1$,

$$\begin{aligned}P(S_{N(\delta^\alpha)} \leqslant x, N(\delta^\alpha)=k) &= P(-k\log(\delta^\alpha) + \Gamma_k' \leqslant x)\; P(N(\delta^\alpha)=k)\\ &= P(-N(\delta^\alpha)\log(\delta^\alpha) + \Gamma_{N(\delta^\alpha)}' \leqslant x, N(\delta^\alpha)=k).\end{aligned}$$

The same relation holds for $k=0$ with $\Gamma_0'=0$. We finally conclude that the limit distribution in (4.2) has on $\{N(\delta^\alpha)>0\}$ the same distribution as

$$\frac{1}{\alpha}\frac{-N(\delta^\alpha)\log(\delta^\alpha) + \Gamma_{N(\delta^\alpha)}'}{N(\delta^\alpha)} + \log\delta = \frac{1}{\alpha}\frac{\Gamma_{N(\delta^\alpha)}'}{N(\delta^\alpha)}, \tag{4.3}$$

where we interpret the right-hand expression as zero if $N(\delta^\alpha)=0$.

From (4.3) we conclude that the right-hand side has expectation α^{-1} and variance

$$\alpha^{-2}[P(N(\delta^\alpha)=0) + E[N(\delta^\alpha)I_{\{N(\delta^\alpha)\}>0}]^{-1}].$$

As $\delta \to 0$, the variance is of the order $\alpha^{-2}[1+\delta^\alpha]$.

4.2. Poisson convergence of the K-function

In the spatial analysis of point patterns the K-function is used as a measure of spatial dependence; see Cressie (1993). A sample version of it is given by the U-statistic

$$K_n(\delta) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} I_{[0,\delta]}(a_n\|\mathbf{X}_i - \mathbf{X}_j\|).$$

Thus we are in the framework of Theorem 2.1 for the particular kernel

$$h(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|,$$

and so we may conclude that $K_n(\delta) = N_n(\mathbf{E}_1 \times [0,\delta])$ converges in distribution to a Poisson random variable with mean δ^α. More generally, the K_n-processes converge in distribution in $M_p(\mathbb{R}_+)$ to a Poisson process K with mean measure $\alpha x^{\alpha-1}\,\mathrm{d}x$:

$$K_n(\cdot) = N_n(\mathbf{E}_1 \times \cdot) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} \varepsilon_{a_n|\mathbf{X}_i - \mathbf{X}_j|}(\cdot) \xrightarrow{\mathrm{d}} K(\cdot). \tag{4.4}$$

Writing $K(\delta)=K([0,\delta])$, it follows from (2.11) that the cumulative processes converge in $\mathbf{D}[0,\infty)$ equipped with the J_1-topology:

$$(K_n(\delta))_{\delta \geqslant 0} \xrightarrow{\mathrm{d}} (K(\delta))_{\delta \geqslant 0}.$$

An application of the continuous mapping theorem on $\mathbf{D}[M_0, M_1]$ with $0 < M_0 < M_1 < \infty$ to $\mathbf{C}[0,b]$ with $b > 0$ gives

$$\begin{aligned} B_n &= \left(\int_{M_0}^{M_1} (\log^+ K_n(\delta) - \beta \log \delta)^2 \,\mathrm{d}\delta \right)_{\beta \in [0,b]} \\ &\xrightarrow{\mathrm{d}} B = \left(\int_{M_0}^{M_1} (\log^+ K(\delta) - \beta \log \delta)^2 \,\mathrm{d}\delta \right)_{\beta \in [0,b]}, \end{aligned}$$

in $\mathbf{C}[0,b]$. Another application of the continuous mapping shows that the minimizer β_n of B_n on $[0,b]$ converges to the minimizer β_0 of B on $[0,b]$:

$$\begin{aligned} \beta_n &= \frac{\int_{M_0}^{M_1} \log \delta \log^+ K_n(\delta)\,\mathrm{d}\delta}{\int_{M_0}^{M_1} (\log \delta)^2\,\mathrm{d}\delta} \\ &\xrightarrow{\mathrm{d}} \beta_0 = \frac{\int_{M_0}^{M_1} \log \delta \log^+ K(\delta)\,\mathrm{d}\delta}{\int_{M_0}^{M_1} (\log \delta)^2\,\mathrm{d}\delta} \\ &= \alpha + \frac{\int_{M_0}^{M_1} (\log \delta \log^+ K(\delta) - \log(\delta^\alpha))\,\mathrm{d}\delta}{\int_{M_0}^{M_1} (\log \delta)^2\,\mathrm{d}\delta}, \end{aligned} \tag{4.5}$$

where $\log^+ x = \log(\max(1,x))$. It seems hard to evaluate the distribution of the integral in (4.5) explicitly. Using a simple simulation, we estimated the bias term in (4.5). This is shown in Fig. 1.

4.3. Hill estimation of α

In extreme value theory the estimation of α is usually not based on log regression methods but on an increasing number of logarithms of lower order statistics. In this section we show that similar techniques work for estimating α when assumption (1.3) is satisfied. We restrict ourselves to one semi-parametric estimator of α, the Hill estimator. The properties of the latter cannot be studied in the context of the point processes N_n defined in (1.4). The latter point processes are constructed in such a way that any bounded set of the state space contains only a finite number of the points of the processes. In order to make the Hill estimator work one needs an increasing number

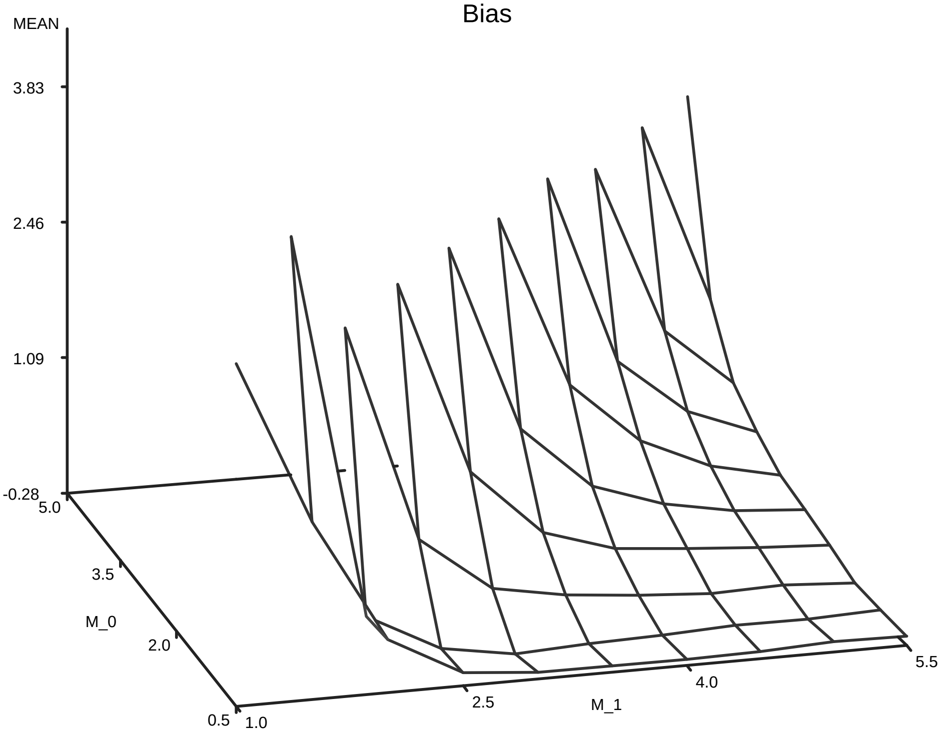

Fig. 1. The vertical axis represents the bias in estimating α, i.e. the fraction in line (4.5). The horizontal axes represent possible values for M_0 and M_1 in that fraction.

of points, i.e. order statistics of the $h(\mathbf{X}_i, \mathbf{X}_j)$'s, in any neighborhood of the origin. The right tool in this context is the tail empirical process as used in the proof of Proposition 4.1 below.

As before, write

$$h_{(1)} \leqslant \cdots \leqslant h_{(n(n-1)/2)}$$

for the order statistics of the sample $h(\mathbf{X}_i, \mathbf{X}_j)$, $i = 2, \ldots, n$, $j = 1, \ldots, i-1$. A classical estimator of α is *Hill's estimator* given by

$$\hat{\alpha}_{n,m} = -\left(\frac{1}{m}\sum_{i=1}^{m}\log(h_{(i)}/h_{(m)})\right)^{-1}$$

for $m \geqslant 1$; see Hill (1975); cf. Embrechts et al. (1997, Chapter 6), for a discussion of the properties of this estimator based on the order statistics of iid and stationary sequences.

Proposition 4.1. *Assume* (1.3) *and* (2.1) *hold. If* $m = m_n \to \infty$ *and* $\sqrt{m_n}/n \to 0$, *then Hill's estimator is consistent, i.e.* $\hat{\alpha}_{n,m} \xrightarrow{P} \alpha$.

Proof. We follow the approach of Resnick and Stărică (1995). Following the lines of their Section 2, it suffices to prove that the tail empirical process $N_{n,m}$ of the points $h(\mathbf{X}_i, \mathbf{X}_j)$ satisfies

$$N_{n,m} = \frac{1}{m} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \varepsilon_{a_{n,m} h(\mathbf{X}_i, \mathbf{X}_j)} \xrightarrow{\mathrm{d}} \mu,$$

in $M_{\mathrm{p}}(\mathbb{R}_+)$, where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution, μ is a measure on the Borel sets of $\mathbb{R}_+$ with density $0.5\alpha x^{\alpha-1}\,\mathrm{d}x$ and $a_{n,m} \to \infty$ is chosen such that

$$\frac{n^2}{2m} P(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2) \leqslant 1) \sim 1. \tag{4.6}$$

Since μ is deterministic it suffices to show that the law of large numbers

$$N_{n,m}((a,b]) \xrightarrow{\mathrm{P}} \mu((a,b])$$

holds for any $0 < a < b < \infty$. Thus it suffices to verify that

$$EN_{n,m}((a,b]) \to \mu((a,b]) \quad \text{and} \quad \mathrm{var}(N_{n,m}((a,b])) \to 0.$$

By the definition of $(a_{n,m})$ and by regular variation we have

$$EN_{n,m}((a,b]) = \frac{n(n-1)}{2m} P(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2) \in (a,b]) \to \mu((a,b]).$$

Applying a standard result for the variance of U-statistics (cf. Lee, 1990, p. 12), we have

$$\mathrm{var}(N_{n,m}((a,b])) \sim \frac{1}{m^2} \left[\frac{n^2}{2} \mathrm{var}(I_{(a,b]}(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2))) \right.$$
$$\left. + \, n^3 \operatorname{cov}(I_{(a,b]}(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2), I_{(a,b]}(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_3)) \right].$$

By virtue of (4.6) it is easily seen that the first term on the right-hand side converges to 0. The covariance on the right-hand side can be written as

$$\frac{n^3}{m^2} [P(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2) \in (a,b], a_{n,m} h(\mathbf{X}_1, \mathbf{X}_3) \in (a,b]) - (P(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2) \in (a,b]))^2]$$
$$= \frac{n^3}{m^2} P(a_{n,m} h(\mathbf{X}_1, \mathbf{X}_2) \in (a,b], a_{n,m} h(\mathbf{X}_1, \mathbf{X}_3) \in (a,b]) + \mathrm{o}(1).$$

In the last step we again used (4.6). However, the first term on the right-hand side also converges to 0 by virtue of (2.1). □

Acknowledgements

The authors are indebted to the referee whose detailed suggestions led to a substantial improvement of our presentation. Part of this research was done when HGD and TM worked at the Department of Mathematics is Groningen and/or EURANDOM Eindhoven. Financial support from Groningen for the visitors ARD and OS is gratefully acknowledged. The support of ARD by the Natural Sciences and Engineering Research Council of Canada is also gratefully acknowledged.

References

Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., Weiss, B., 1996. Strong laws for L- and U-statistics. Trans. Amer. Math. Soc. 348, 2845–2866.

Barbour, A.D., Eagleson, G.K., 1984. Poisson convergence for dissociated statistics. J. Roy. Statist. Soc. Ser. B 46, 397–402.

Barbour, A.D., Holst, L., Janson, S., 1992. Poisson Approximation. Oxford University Press, New York.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Cambridge University Press, Cambridge.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Daley, D., Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer Series in Statistics. Springer, New York.

Davis, R.A., Hsing, T., 1995. Point process and partial sum convergence for weakly dependent random variables with infinite variance. Ann. Probab. 23, 879–917.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Grassberger, P., Procaccia, I., 1983. Characterization of strange attractors. Phys. Rev. Lett. 50, 346–349.

Heinrich, L., Wolf, W., 1993. On the convergence of U-statistics with stable limit distribution. J. Multivariate Anal. 44, 266–278.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.

Kallenberg, O., 1983. Random Measures, 3rd Edition. Akademie-Verlag, Berlin.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer, Berlin.

Lee, A.J., 1990. U-Statistics. Theory and Practice. Marcel Dekker, New York.

Resnick, S.I., 1986. Point processes, regular variation and weak convergence. Adv. Appl. Probab. 18, 66–138.

Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer, New York.

Resnick, S.I., Stă rică , C., 1995. Consistency of Hill's estimator for dependent data. J. Appl. Probab. 32, 139–167.

Samorodnitsky, G., Taqqu, M.S., 1994. Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance. Chapman and Hall, London.

Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Silverman, B., Brown, T., 1978. Short distances, flat triangles and Poisson limits. J. Appl. Probab. 15, 815–825.

Takens, F., 1985. On the numerical determination of the dimension of an attractor. In: Dynamical Systems and Bifurcations (Groningen, 1984), Lecture Notes in Mathematics, Vol. 1125, Springer, Berlin, pp. 99–106.